

Universidad Nacional de Colombia

Clasificador de música por géneros

Introducción a la Inteligencia Artificial

Paola Andrea Gallegos Pinto - 1005257935 - pgallegos@unal.edu.co
Vanessa Lozano - 1000795707 - vlozano@unal.edu.co

3 de diciembre de 2022

Índice

1. Introducción	1
1.1. El problema:	1
1.2. ¿Qué es la clasificación de géneros musicales?	2
1.3. ¿Por qué es importante la clasificación de géneros musicales?	2
1.4. ¿Cómo clasificar los géneros musicales?	2
2. Materiales y métodos	2
2.1. Materiales:	2
2.2. Métodos:	4
2.2.1. KNN:	4
2.2.2. CNN:	5
3. Resultados	6
3.1. KNN:	6
3.2. CNN:	6
4. Discusión y conclusiones	7

1. Introducción

Trabajar con datos de audio ha sido un problema relativamente menos explorado en el machine learning. En la mayoría de los casos, los puntos de referencia para el trabajo fundamental en Deep learning se miden en el rendimiento de datos de texto e imágenes. En medio de esto, el habla y el audio, un tipo de datos igualmente importante, a menudo se pasan por alto. Desde la conversión de palabras habladas a texto, traducción en vivo, el uso de audio y música para análisis o filtrado, los datos de audio tienen una amplia gama de aplicaciones en la vida diaria.

1.1. El problema:

El objetivo de este proyecto es construir modelos de aprendizaje automático que procese archivos de audio (en formato wav para este caso particular) y los clasifique en el género musical correcto.

Diferentes tipos de soluciones y algoritmos de aprendizaje automático son capaces de resolver este problema, desde algoritmos de aprendizaje automático simple y supervisado que se puede usar para resolver problemas de clasificación y regresión como lo es KNN, hasta redes neuronales convolucionales (CNN). Para este proyecto

en particular, nos enfocamos en construir una CNN y KNN que pueda predecir con precisión el género de diferentes archivos de audio .

1.2. ¿Qué es la clasificación de géneros musicales?

Clasificar la música en varios géneros (hip hop, rock, jazz, folk, pop, etc.) implica extraer características valiosas de los datos de audio, preprocesarlos y entrenar un modelo clasificador de aprendizaje automático. Tenemos varias formas efectivas de representar imágenes y textos en forma numérica. Aún así, esta tarea se vuelve más desafiante para los datos de audio, ya que hay varios atributos que podemos considerar para “describir” las señales de audio. Existen diferentes formas de convertir datos de audio en numéricos: utilizando componentes de frecuencia, características acústicas o espectrogramas.

1.3. ¿Por qué es importante la clasificación de géneros musicales?

Por supuesto, la clasificación de audio generalizada tiene aplicaciones de mayor alcance, como la detección de voz, la separación del ruido de fondo del habla principal, el análisis de sentimientos, la identificación del hablante para la verificación de voz. Para empezar, desde el inicio de los servicios de transmisión de música, hemos tenido el lujo de buscar y elegir entre una variedad de música, con sugerencias relevantes después de cada canción escuchada.

Hoy en día, servicios como Spotify y YouTube crean rápidamente listas de reproducción y sugieren la siguiente canción en función de nuestros patrones de escucha. Estos patrones están influenciados por el tono, el estado de ánimo o el género de las canciones que escuchamos. Varias variables influyen en la decisión, pero la mayoría tiene que ver con cómo “suena” la canción. Por ejemplo, una canción pop puede sentirse más rápida y “funky” que, digamos, una canción romántica; Con base en parámetros específicos que definen estas características, los servicios de transmisión de música agrupan canciones con sonidos o sentimientos similares en el mismo género. Por lo tanto, clasificar la música en función de los géneros puede ayudar a sugerir las siguientes canciones a un oyente, seleccionar listas de reproducción de nuevas recomendaciones o filtrar contenido no deseado.

1.4. ¿Cómo clasificar los géneros musicales?

Seleccionar la representación de datos es más importante antes de elegir entre una gran cantidad de algoritmos de aprendizaje automático disponibles para la clasificación. La conversión de datos de audio a formato numérico o vectorial determinará cuánta información fundamental se retiene cuando se pierde la forma de audio. Por ejemplo, si un formato de datos no puede representar el volumen y el ritmo de una canción de rock, sería difícil incluso para los mejores modelos de aprendizaje automático aprender el género y clasificar la muestra.

De los diversos métodos para representar datos de audio, algunos son los siguientes: los espectrogramas de Mel visualizan señales de audio en función de sus componentes de frecuencia, mientras que algunos conjuntos de datos (ejemplo: Problema de clasificación de Spotify 2021) usan otras especificaciones como BPM, volumen, acústica, “energía”, “vivacidad”, etc.

2. Materiales y métodos

2.1. Materiales:

Para cualquier proyecto de machine learning, hay dos cosas principales: extracción de características de los datos; y entrenar al modelo. Para la extracción de características de audio y música con fines de aprendizaje automático, por lo general, los coeficientes cepstrales de frecuencia mel (MFCC) se extraen de la canción o el audio, para este proyecto hicimos uso de la base de datos libre GTZAN[2] que contiene 10 géneros musicales: blues, clásica, country, disco, hiphop, jazz, metal, pop, reggae y rock. Para cada uno de estos se tienen 100 muestras de audio de 30 segundos, cada una con su representación visual (espectrograma). Contiene

dos archivos .csv el primero nos proporciona 58 características acústicas para cada uno de los audios de 30 segundos y el segundo archivo con el análisis de estas mismas características pero para fragmentos de 3 segundos de cada audio, incrementando 10 veces el tamaño de los datos. Estas características se utilizan para entrenar el modelo.

Es necesario decidir de que forma se tratará esta información como valores numéricos, garantizando que sea eficiente y no se pierdan los atributos distintivos para cada muestra de sonido, como podrían serlo el ritmo o volumen. Las alternativas son: usar la muestra como valores numéricos, comparandolos según las características brindadas por la base de datos, como lo es la variación del MFCC o analizarlos según su espectrograma, es decir a partir de la clasificación de imágenes.

Para explicar esto mejor, cuando representamos un archivo de audio en formato digital, la computadora lo ve como una onda con el eje X como tiempo y el eje Y como amplitud. Esto se muestra en la Figura 1.

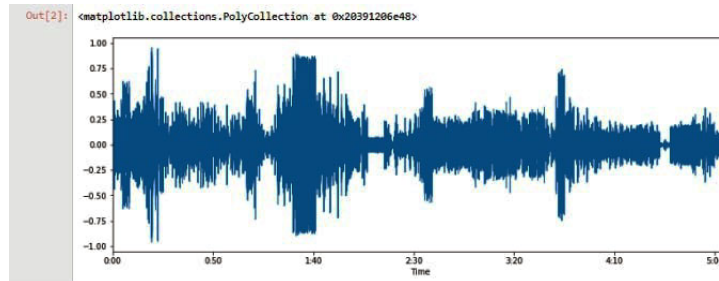


Figura 1: Music representation in amplitude and time

Este formato de representación no nos da mucha información sobre el audio o la canción. Por lo tanto, representamos el audio en el dominio de la frecuencia usando algo llamado Transformada Rápida de Fourier (FFT). Usando esta FFT, convertimos nuestro archivo de audio de entrada y lo representamos en el dominio de la frecuencia y el tiempo. El gráfico que muestra los datos de audio en los dominios de frecuencia y tiempo se denomina espectrograma y se representa en la Figura 2.

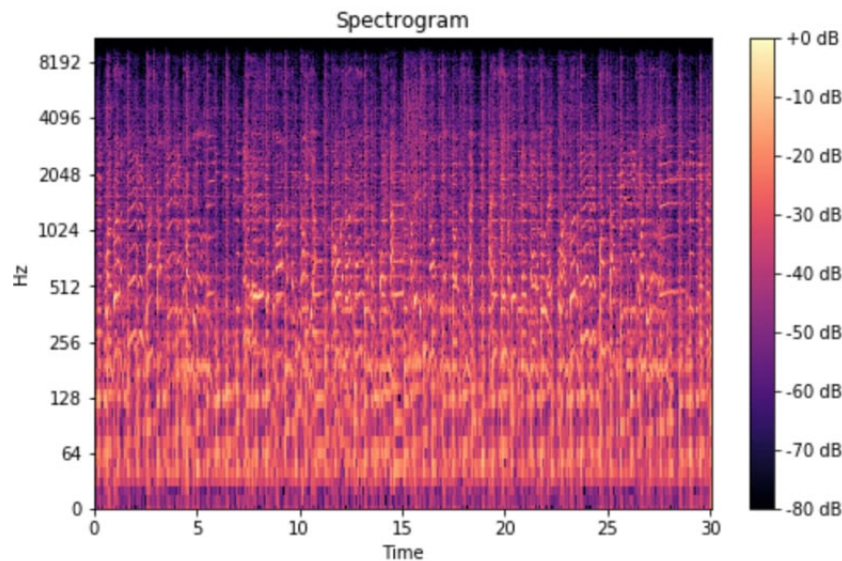


Figura 2: Mel spectrogram

los MFCC utilizan una escala de mel que se usa para extraer las características de una señal de audio, que cuando se representa como un gráfico, resulta ser un espectrograma de mel. Entonces, en pocas palabras, lo

que vemos en un espectrograma de mel son las características exactas que necesitamos para entrenar nuestro modelo.

Ahora bien, dependiendo de como se deseen tratar los datos existen diferentes modelos usados en el campo de la inteligencia artificial para resolver este problema, como lo son [3]:

- **CNN (Convolutional neural network):** Una red neuronal convolucional (ConvNet/CNN) es un algoritmo de aprendizaje profundo que puede tomar una imagen de entrada, asignar importancia (pesos) a varias características en la imagen y poder diferenciar una de otra. CNN trabaja con convolución y agrupación en secuencia alterna en la red neuronal. Entonces, ¿por qué usar CNN en la clasificación de géneros? CNN, como se discutió antes, está especializado en el reconocimiento de características de imagen, así pues como mencionamos anteriormente el espectrograma que representa el MFCC junto con el tiempo y la frecuencia será la entrada del modelo CNN.
- **LSTM (Long short-term memory):** Es un tipo especial de red neuronal recurrente que es capaz de aprender dependencias a largo plazo en los datos. Esto se logra porque el módulo recurrente del modelo tiene una combinación de cuatro capas que interactúan entre sí. Los LSTM son bastante buenos para extraer patrones en el espacio de características de entrada, donde los datos de entrada abarcan secuencias largas. Dada la arquitectura cerrada de los LSTM que tienen esta capacidad de manipular su estado de memoria, son ideales para este problema.
- **KNN (K nearest neighbors):** K nearest neighbors (KNN) es un tipo de algoritmo de aprendizaje supervisado que se utiliza tanto para la regresión como para la clasificación. KNN intenta predecir la clase correcta para los datos de prueba calculando la distancia entre los datos de prueba y todos los puntos de entrenamiento. El algoritmo KNN, cuando se implementa en la clasificación de géneros musicales, analiza canciones similares y asume que pertenecen a la misma categoría porque parecen estar cerca unas de otras. Entre varias otras técnicas que prevalecen en este concepto, los mejores resultados se han obtenido de esta técnica.

2.2. Métodos:

Decidimos implementar los modelos KNN y CNN, para la clasificación a partir de datos y espectrogramas respectivamente, para poder comparar la precisión de ambos métodos.

2.2.1. KNN:

Para usar el modelo de los K vecinos más cercanos dividimos el proceso en 3 partes ilustradas en los siguientes diagramas de bloques:

- **Procesamiento de la base de datos:** Para usar nuestra base de datos GTZAN debemos extraer para cada audio sus características, que son datos de tipo time-series, es decir, son datos extensos, por lo que decidimos guardar esta información en dos archivos binarios haciendo una separación aleatoria de datos para entrenar el modelo o testarlo.

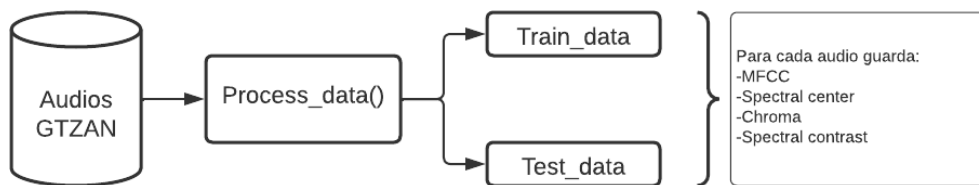


Figura 3: Procesamiento de datos KNN y CNN

- **Calcular precisión del modelo:** Comparamos la distancia o similitud de las características de cada caso de prueba con los elementos de train data para hallar los K elementos más cercanos a nuestra prueba

(para cada característica). Luego se busca el género más cercano en todas las características y el modelo predice cual es el género musical. Comparamos estas predicciones con los géneros correctos del test data y así obtenemos la precisión de nuestro modelo.

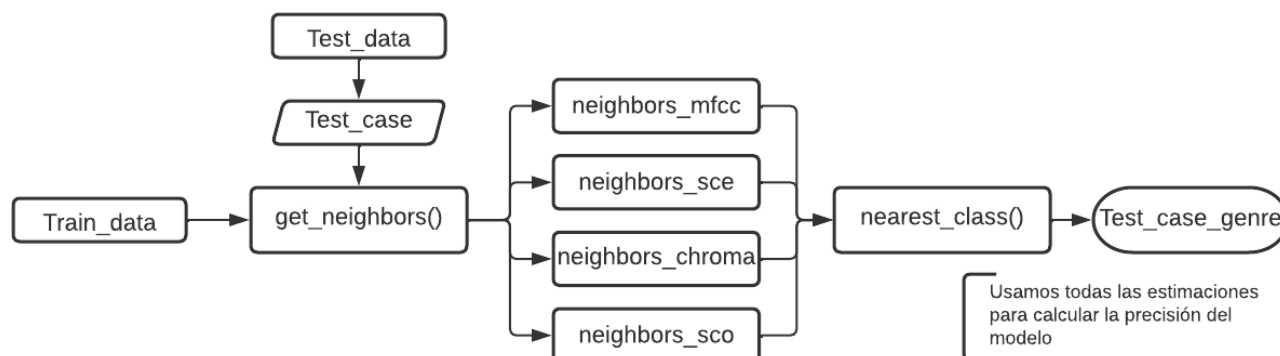


Figura 4: Entrenamiento y cálculo de precisión KNN

- **Funcionamiento:** El usuario puede escoger un audio para analizar. De igual forma que en el entrenamiento, se decide el género en base a la distancia o similitud del audio nuevo con los datos de train data a partir de sus vecinos más cercanos

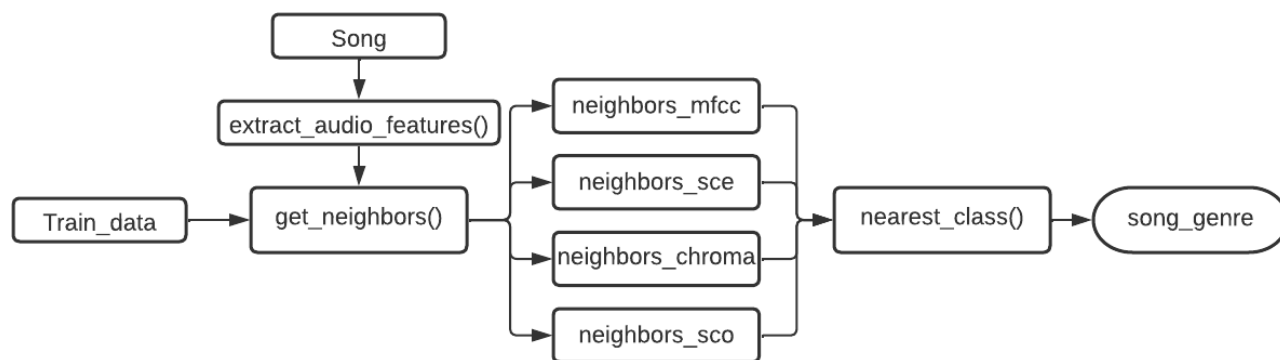


Figura 5: Funcionamiento KNN

2.2.2. CNN:

Para usar el modelo de los CNN dividimos el proceso en 3 partes ilustradas en los siguientes diagramas de bloques:

- **Procesamiento de la base de datos:** Para usar nuestra base de datos GTZAN debemos extraer para cada audio sus características en un archivo .json en cual contendra los MFCC del conjunto de datos de música de cada genero musical y sus canciones junto con las etiquetas de género. Usando librosa, se pueden extraer características significativas de los datos de audio, como los coeficientes cepstrales de frecuencia Mel (MFCC), el centroide espectral, etc. dominio, para presentar la frecuencia en una escala que es más perceptible para el oído humano. Luego calcula un espectrograma de esas nuevas frecuencias escaladas, lo que produce representaciones visuales del espectro de frecuencias y su intensidad.

- Calcular precisión del modelo: Este fue nuestro modelo más avanzado, utilizando 3 capas de convolución, cada una con su propio grupo máximo y regularización, alimentación en 3 capas totalmente conectadas con activación ReLU, salida softmax y pérdida de entropía cruzada. La mayoría de las ecuaciones pueden se encuentra arriba, este enfoque involucra ventanas de convolución que escanean los datos de entrada y generan la suma de los elementos dentro del ventana. Esto luego se alimenta a una capa de grupo máximo que selecciona el elemento máximo de otra ventana. Esto se implementó con TensorFlow y Keras [4] [5].

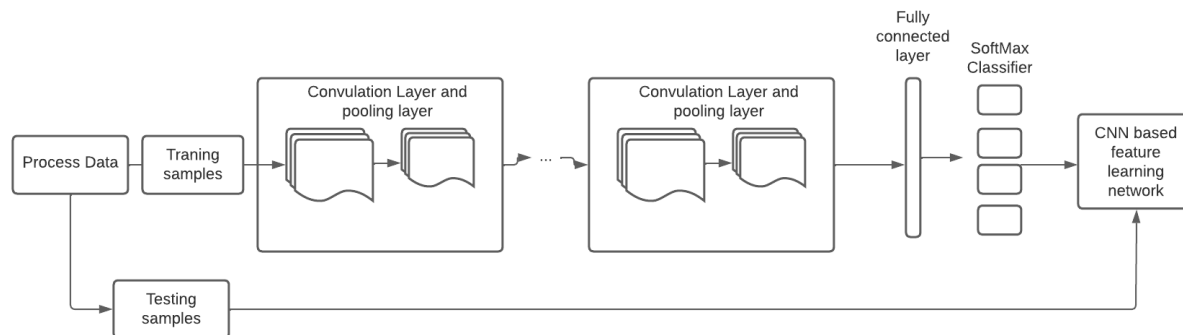


Figura 6: Creacion y Entrenamiento CNN

- Funcionamiento: El usuario puede escoger un audio para analizar. De igual forma que en el entrenamiento, se decide el género utilizando 3 capas de convolución, cada una con su propio grupo máximo y regularización, alimentación en 3 capas totalmente conectadas con activación ReLU.

3. Resultados

3.1. KNN:

A partir de varios experimentos realizados pudimos determinar lo siguiente:

- La efectividad del modelo depende del K escogido pero oscila entre un 40-60 %, para k muy grande oscila entre 10-20 %
- La efectividad del modelo depende de la medida de distancia que se utilice entre las características y la precisión de esta
- Es más probable que el modelo se equivoque escogiendo entre algunos géneros que pueden tener características similares, por ejemplo rock y metal. Esto podría explicarse por la distribución de los géneros de música en el espacio, como ilustra el siguiente enlace: <https://everynoise.com/> , entre más cercanos sean dos géneros, más probable va a ser que el modelo se equivoque.

En el Cuadro 1 podemos observar algunos ejemplos

3.2. CNN:

A partir de varios experimentos realizados pudimos determinar lo siguiente:

- La efectividad del modelo oscila entre un 70-75 % con 50 y 80 epoch
- CNN tuvo más problemas con el género rock. Sólo logró clasificar correctamente el 50 % del audio de rock como rock, etiquetando a los demás como principalmente country o blues. Además, clasificó incorrectamente algunos Country, así como una pequeña fracción de blues y reggae, como música rock. Si bien no es tan sorprendente que el rock fuera un género desafiante: una inspección cualitativa de

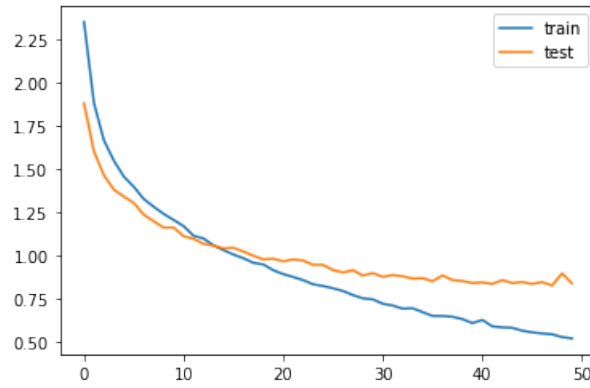


Figura 7: loss: 0.8306 - accuracy: 0.7101 - 1s/epoch - 18ms/step

los espectrogramas de rock mel implica que muchos extractos de música rock carecen de los ritmos fácilmente visibles que poseen otros géneros como el hip-hop y el disco.

- la música rock frente a otros géneros nos dice que al rock también le faltan rasgos distintivos como voces de registro alto (pop) o piano fácilmente audible (clásica o jazz). Además, el rock es un género que encapsula muchos estilos diferentes (rock ligero, rock duro, rock progresivo, rock independiente, new wave, etc.) e influye fuertemente en muchos otros géneros derivados. Sin embargo, nos sorprendió que el rock y el country se confundieran tan fácilmente, a diferencia del rock y metal, que parece depender de una instrumentación y un tempo más similares.

Veamos algunos ejemplos:

Ejemplos de 30 segundos MP3

Cancion	Genero esperado	Genero
Back in black - ACDC	Rock	Country
Highway to Hell - ACDC	Rock	Rock
I fell in love to easily - Chet Baker	Jazz	Blues
I will survive - Gloria Gaynor	Disco	Disco
Is this love - Bob Marley	Reggae	Reggae
Cant make you love me -Britney Spears	Pop	Pop

Figura 8: Ejemplos canciones mp3 de 30 segundos

4. Discusión y conclusiones

Conclusiones:

- El tamaño de la base de datos y los recursos computacionales con los que contabamos, no son suficientes para construir un modelo con la precisión necesaria para resolver este problema, aunque logramos un acercamiento satisfactorio

- El uso de redes neuronales convolucionales CNN aplicado a los espectrogramas, presentó mejor desempeño en comparación al modelo de K vecinos más cercanos
- De los modelos, el uso de espectrogramas de mel basados en frecuencia produjo resultados de mayor precisión. Además, los espectrogramas mel son visuales y las CNN funcionan mejor con imágenes. la cnn se desempeñó mejor, como esperábamos. También tomó más tiempo entrenar, pero el aumento en la precisión justifica la costo adicional de cómputo.
- En el futuro, esperamos experimentar con otros tipos de métodos de aprendizaje profundo, dado que se desempeñaron mejor. Dado que se trata de datos de series temporales, algún tipo de modelo RNN puede funcionar bien (GRU, LSTM, por ejemplo). Adentraros en el tema sobre los aspectos generativos de este proyecto, incluyendo algún tipo de conversión de género (en la misma línea que generativo redes antagónicas que pintan fotos al estilo de Van Gogh, pero específicamente para la música).

Referencias

- [1] Bosch research, "SoundSee – insight with Audio AI"[Online]. Disponible en: <https://www.bosch.com/research/know-how/success-stories/audio-ai/>
- [2] Andrada Olteanu (2019), "GTZAN Dataset - Music Genre Classification"[Online]. Disponible en: <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification?resource=download>
- [3] Data plair (2019), "11 Top Machine Learning Algorithms used by Data Scientists"[Online]. Disponible en: <https://data-flair.training/blogs/machine-learning-algorithms/>
- [4] François Chollet et al. Keras. Disponible en: <https://keras.io> , 2015.
- [5] Large-scale machine learning on heterogeneous systems, 2015. Software Disponible en: tensorflow.org.

Canción	Género esperado	Resultado	Distribución																																																							
AC/DC- Highway to Hell	Rock	Rock	<table><caption>K neighbors distribution</caption><thead><tr><th>Genre</th><th>MFCC</th><th>Spectral center</th><th>Chroma</th><th>Spectral contrast</th></tr></thead><tbody><tr><td>blues</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>classical</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>country</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>disco</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>hiphop</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>jazz</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>metal</td><td>0.10</td><td>0.10</td><td>0.15</td><td>0.30</td></tr><tr><td>pop</td><td>0.15</td><td>0.20</td><td>0.20</td><td>0.05</td></tr><tr><td>reggae</td><td>0.10</td><td>0.15</td><td>0.05</td><td>0.10</td></tr><tr><td>rock</td><td>0.65</td><td>0.55</td><td>0.60</td><td>0.55</td></tr></tbody></table>	Genre	MFCC	Spectral center	Chroma	Spectral contrast	blues	0.00	0.00	0.00	0.00	classical	0.00	0.00	0.00	0.00	country	0.00	0.00	0.00	0.00	disco	0.00	0.00	0.00	0.00	hiphop	0.00	0.00	0.00	0.00	jazz	0.00	0.00	0.00	0.00	metal	0.10	0.10	0.15	0.30	pop	0.15	0.20	0.20	0.05	reggae	0.10	0.15	0.05	0.10	rock	0.65	0.55	0.60	0.55
Genre	MFCC	Spectral center	Chroma	Spectral contrast																																																						
blues	0.00	0.00	0.00	0.00																																																						
classical	0.00	0.00	0.00	0.00																																																						
country	0.00	0.00	0.00	0.00																																																						
disco	0.00	0.00	0.00	0.00																																																						
hiphop	0.00	0.00	0.00	0.00																																																						
jazz	0.00	0.00	0.00	0.00																																																						
metal	0.10	0.10	0.15	0.30																																																						
pop	0.15	0.20	0.20	0.05																																																						
reggae	0.10	0.15	0.05	0.10																																																						
rock	0.65	0.55	0.60	0.55																																																						
Chopin Nocturne op.9 No.2	Classical	Classical	<table><caption>K neighbors distribution</caption><thead><tr><th>Genre</th><th>MFCC</th><th>Spectral center</th><th>Chroma</th><th>Spectral contrast</th></tr></thead><tbody><tr><td>blues</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>classical</td><td>0.40</td><td>0.80</td><td>0.60</td><td>0.20</td></tr><tr><td>country</td><td>0.20</td><td>0.20</td><td>0.20</td><td>0.20</td></tr><tr><td>disco</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>hiphop</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>jazz</td><td>0.20</td><td>0.20</td><td>0.20</td><td>0.40</td></tr><tr><td>metal</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>pop</td><td>0.20</td><td>0.20</td><td>0.20</td><td>0.20</td></tr><tr><td>reggae</td><td>0.00</td><td>0.20</td><td>0.00</td><td>0.00</td></tr><tr><td>rock</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr></tbody></table>	Genre	MFCC	Spectral center	Chroma	Spectral contrast	blues	0.00	0.00	0.00	0.00	classical	0.40	0.80	0.60	0.20	country	0.20	0.20	0.20	0.20	disco	0.00	0.00	0.00	0.00	hiphop	0.00	0.00	0.00	0.00	jazz	0.20	0.20	0.20	0.40	metal	0.00	0.00	0.00	0.00	pop	0.20	0.20	0.20	0.20	reggae	0.00	0.20	0.00	0.00	rock	0.00	0.00	0.00	0.00
Genre	MFCC	Spectral center	Chroma	Spectral contrast																																																						
blues	0.00	0.00	0.00	0.00																																																						
classical	0.40	0.80	0.60	0.20																																																						
country	0.20	0.20	0.20	0.20																																																						
disco	0.00	0.00	0.00	0.00																																																						
hiphop	0.00	0.00	0.00	0.00																																																						
jazz	0.20	0.20	0.20	0.40																																																						
metal	0.00	0.00	0.00	0.00																																																						
pop	0.20	0.20	0.20	0.20																																																						
reggae	0.00	0.20	0.00	0.00																																																						
rock	0.00	0.00	0.00	0.00																																																						
Olivia Rodri- go - drivers license	Pop	Pop	<table><caption>K neighbors distribution</caption><thead><tr><th>Genre</th><th>MFCC</th><th>Spectral center</th><th>Chroma</th><th>Spectral contrast</th></tr></thead><tbody><tr><td>blues</td><td>0.00</td><td>0.00</td><td>0.20</td><td>0.00</td></tr><tr><td>classical</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.40</td></tr><tr><td>country</td><td>0.20</td><td>0.00</td><td>0.00</td><td>0.20</td></tr><tr><td>disco</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>hiphop</td><td>0.00</td><td>0.20</td><td>0.00</td><td>0.00</td></tr><tr><td>jazz</td><td>0.00</td><td>0.20</td><td>0.00</td><td>0.00</td></tr><tr><td>metal</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>pop</td><td>0.40</td><td>0.40</td><td>0.40</td><td>0.20</td></tr><tr><td>reggae</td><td>0.40</td><td>0.20</td><td>0.20</td><td>0.00</td></tr><tr><td>rock</td><td>0.00</td><td>0.00</td><td>0.20</td><td>0.20</td></tr></tbody></table>	Genre	MFCC	Spectral center	Chroma	Spectral contrast	blues	0.00	0.00	0.20	0.00	classical	0.00	0.00	0.00	0.40	country	0.20	0.00	0.00	0.20	disco	0.00	0.00	0.00	0.00	hiphop	0.00	0.20	0.00	0.00	jazz	0.00	0.20	0.00	0.00	metal	0.00	0.00	0.00	0.00	pop	0.40	0.40	0.40	0.20	reggae	0.40	0.20	0.20	0.00	rock	0.00	0.00	0.20	0.20
Genre	MFCC	Spectral center	Chroma	Spectral contrast																																																						
blues	0.00	0.00	0.20	0.00																																																						
classical	0.00	0.00	0.00	0.40																																																						
country	0.20	0.00	0.00	0.20																																																						
disco	0.00	0.00	0.00	0.00																																																						
hiphop	0.00	0.20	0.00	0.00																																																						
jazz	0.00	0.20	0.00	0.00																																																						
metal	0.00	0.00	0.00	0.00																																																						
pop	0.40	0.40	0.40	0.20																																																						
reggae	0.40	0.20	0.20	0.00																																																						
rock	0.00	0.00	0.20	0.20																																																						
Bob Marley - Is this love	Reggae	Reggae- Rock	<table><caption>K neighbors distribution</caption><thead><tr><th>Genre</th><th>MFCC</th><th>Spectral center</th><th>Chroma</th><th>Spectral contrast</th></tr></thead><tbody><tr><td>blues</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>classical</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>country</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>disco</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.40</td></tr><tr><td>hiphop</td><td>0.00</td><td>0.20</td><td>0.00</td><td>0.00</td></tr><tr><td>jazz</td><td>0.00</td><td>0.00</td><td>0.00</td><td>0.00</td></tr><tr><td>metal</td><td>0.00</td><td>0.20</td><td>0.40</td><td>0.00</td></tr><tr><td>pop</td><td>0.20</td><td>0.20</td><td>0.20</td><td>0.20</td></tr><tr><td>reggae</td><td>0.40</td><td>0.40</td><td>0.40</td><td>0.20</td></tr><tr><td>rock</td><td>0.40</td><td>0.20</td><td>0.40</td><td>0.20</td></tr></tbody></table>	Genre	MFCC	Spectral center	Chroma	Spectral contrast	blues	0.00	0.00	0.00	0.00	classical	0.00	0.00	0.00	0.00	country	0.00	0.00	0.00	0.00	disco	0.00	0.00	0.00	0.40	hiphop	0.00	0.20	0.00	0.00	jazz	0.00	0.00	0.00	0.00	metal	0.00	0.20	0.40	0.00	pop	0.20	0.20	0.20	0.20	reggae	0.40	0.40	0.40	0.20	rock	0.40	0.20	0.40	0.20
Genre	MFCC	Spectral center	Chroma	Spectral contrast																																																						
blues	0.00	0.00	0.00	0.00																																																						
classical	0.00	0.00	0.00	0.00																																																						
country	0.00	0.00	0.00	0.00																																																						
disco	0.00	0.00	0.00	0.40																																																						
hiphop	0.00	0.20	0.00	0.00																																																						
jazz	0.00	0.00	0.00	0.00																																																						
metal	0.00	0.20	0.40	0.00																																																						
pop	0.20	0.20	0.20	0.20																																																						
reggae	0.40	0.40	0.40	0.20																																																						
rock	0.40	0.20	0.40	0.20																																																						